BMC Bioinformatics

RESEARCH   Open Access

ScaPD: a database for human scaffold proteins

Xiaomei Han[1,2], Jenny Wang[3], Jie Wang[2], Sheng Liu[2], Jianfei Hu[1], Heng Zhu[4,5,6] and Jiang Qian[1,5*]

Abstract

Background: Scaffold proteins play a critical role in an increasing number of biological signaling processes, including simple tethering mechanism, regulating selectivity in pathways, shaping cellular behaviors. While many databases document the signaling pathways, few databases are devoted to the scaffold proteins that medicate signal transduction.

Results: Here, we have developed a user-friendly database, ScaPD, to describe computationally predicted, experimentally validated scaffold proteins and associated signaling pathways. It currently contains 273 scaffold proteins and 1118 associated signaling pathways. The database allows users to search, navigate and download the scaffold protein-mediated signaling networks.

Conclusions: Manually curated and predicted scaffold protein data will be a foundation for further investigation of the scaffold protein in the signal transduction. With maintained up-to-date data, ScaPD (http://bioinfo.wilmer.jhu.edu/ScaPD) will be a valuable resource for understanding how individual signaling pathways are regulated.

Keywords: Scaffold protein, Signaling pathway, Database

Background

About 10% of proteins expressed in human cells are involved in the signal transduction [1]. How can signaling proteins interact with the correct partners and avoid wrong proteins? One principle is that cells achieve well in the signal transduction networks by tethering subset proteins in space and time. More than 20 years ago, the first set of scaffold proteins were discovered, which assemble components of diverse pathways at the plasma membrane or subcellular compartments [2–6]. For example, scaffold protein Ste5 tethers multiple protein kinases in the MAP kinase cascade, such as Ste11, Ste7 and Fus3. The spatial organization achieves high efficacy information transfer on cellular information flow.

The scaffold proteins link multiple signaling proteins together to facilitate signal transduction [6, 7]. These proteins mediate a linear pathway among many partner proteins, and mediate pathway branching to multiple outputs as well [8, 9]. One central role of scaffold proteins is to coordinate feedback loops in signaling pathways, and thus to regulate the signaling response [10, 11]. They enhance signaling specificity or increase the signaling efficiency by increasing the local concentration of signaling components. Thus, the scaffold proteins play a crucial role in the signal transduction.

Although various signaling pathways are the central topics in many biological fields, researchers pay much less attention on the scaffold proteins. One possible reason is that identification of scaffold proteins is challenging, which requires multiple steps using traditional biochemical techniques, including selection of a candidate scaffold protein, testing the protein-protein interaction and assessment of the signaling pathway. The systematic study of scaffold proteins can greatly enhance
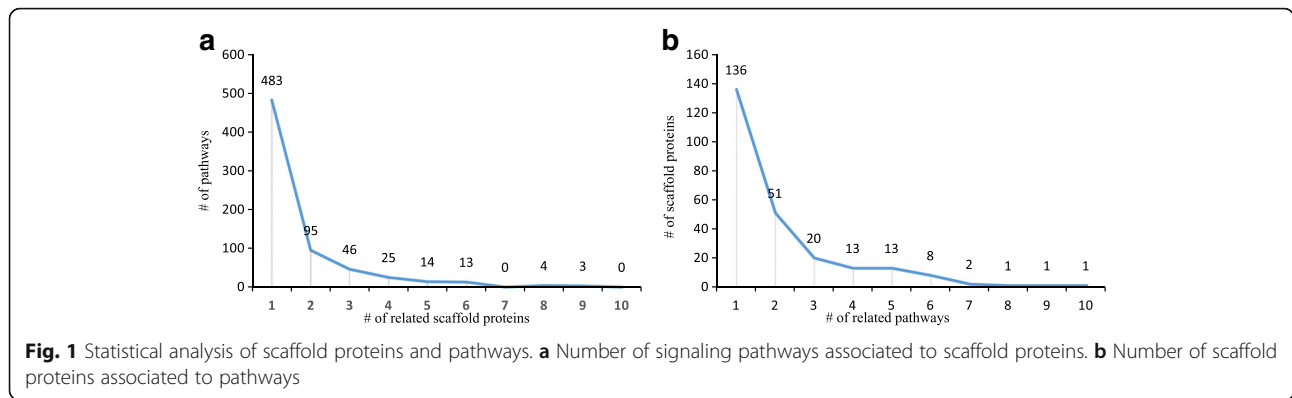
* Correspondence: jiang.qian@jhmi.edu
[1]Department of Ophthalmology, Johns Hopkins School of Medicine, Baltimore, MD, USA
[5]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, USA
Full list of author information is available at the end of the article

Fig. 1 Statistical analysis of scaffold proteins and pathways. **a** Number of signaling pathways associated to scaffold proteins. **b** Number of scaffold proteins associated to pathways

the understanding of the protein regulation that occurs in eukaryotic organisms [12–14].

While many databases were constructed to collect the information of signaling pathways (e.g. KEGG, phospho-networks, phosphoGRID) [15–17], these databases often contain little information of the scaffold proteins. We believe that a central portal specifically designed for scaffold proteins will provide a useful resource to the research community. To facilitate usage of the information of scaffold proteins, we created a scaffold protein database, ScaPD, an integrated information system for the storage and visualization of human scaffold proteins as well as the corresponding signaling pathway data.
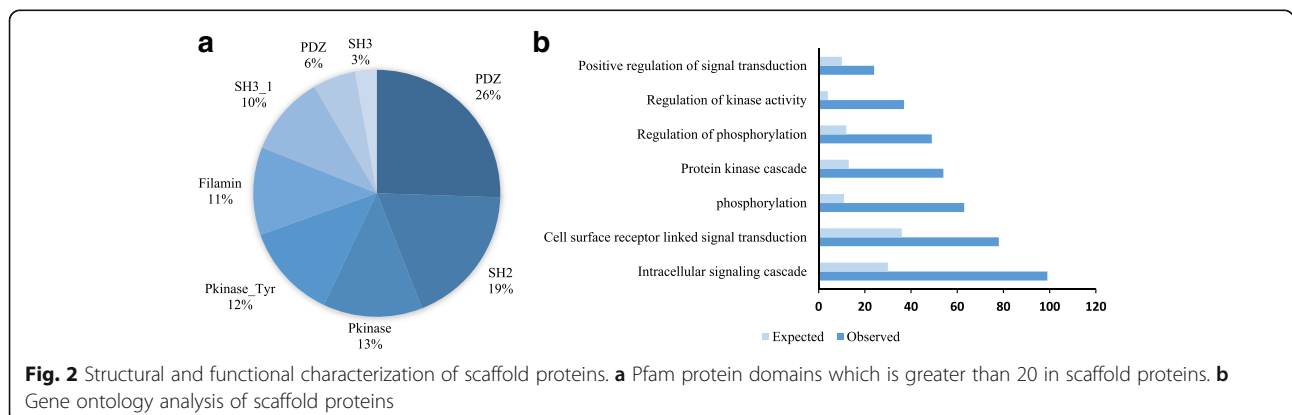
Results

The content of the database has two major sources. First, we performed a manual curation of the literatures to collect experimentally determined scaffold proteins. We first searched papers containing the keyword "scaffold protein" through Google Scholar and PubMed. We then manually examined the papers and collected the known scaffold proteins. In total, we collected 82 scaffold proteins. Second, we collected predicted scaffold proteins generated from a recent project, in which we developed a bioinformatics approach to predict scaffold proteins [18]. In brief, we constructed a composite network, including 55,048 protein-protein interactions and 1103 kinase-substrate relationship in human. We then identified the proteins that interact with multiple components in a signaling pathway. Based on our analysis, 212 proteins were predicted as scaffold proteins with statistical significance.

In total, ScaPD collected 273 scaffold proteins and 683 distinct scaffold-mediated phosphorylation pathways. The association between scaffold proteins and signaling pathways are specific. In fact, 483 (70%) of signaling pathways are associated with only one scaffold protein (Fig. 1a), and 136(51%) of scaffold proteins are associated with one pathway (Fig. 1b).

The scaffold proteins often contain certain protein domains based on Pfam annotation [19]. The most prevalent domains are PDZ (26%), SH2 (19%) and Pkinase domains (13%) (Fig. 2a). The gene ontology (GO) annotation analysis indicates that 99 of the 273 scaffold proteins are associated with "intracellular signal transduction" ($p < 1 \times 10^{-39}$, hypergeometric distribution), and that 75 of predicted scaffold proteins with "phosphorylation" ($p < 1 \times 10^{-23}$, hypergeometric distribution), both over three-fold enrichment than expected group (Fig. 2b).

Users can input any human protein name, and depending on whether the protein of entry is a scaffold protein and/or signaling protein, ScaPD will return a

Fig. 2 Structural and functional characterization of scaffold proteins. **a** Pfam protein domains which is greater than 20 in scaffold proteins. **b** Gene ontology analysis of scaffold proteins

GAB2 is a predicted scaffold protein, which is related to the following pathway(s).

LYN   -> EPOR
LYN   -> PLCG2
MAPK3 -> ETV6
SRC   -> PIK3R1
SRC   -> PLCG2
SRC   -> PTPN6
SRC   -> SHC1
SYK   -> SHC1
MAPK1 -> EGFR   -> GRB2
MAPK1 -> LCK   -> PLCG2
MAPK1 -> LCK   -> PTPN6
MAPK1 -> LCK   -> SHC1
MAPK3 -> EGFR   -> GRB2
MAPK3 -> LCK   -> PLCG2
MAPK3 -> LCK   -> PTPN6
MAPK3 -> LCK   -> SHC1
SRC   -> EGFR   -> GRB2
MAPK1 -> LCK   -> ZAP70 -> SHC1
MAPK3 -> LCK   -> ZAP70 -> SHC1

GAB2 is a known scaffold protein, which is related to the following pathway(s).

SHP2 -> P85

Reference:Scaffolding protein Gab2 mediates differentiation signaling downstream of Fms receptor tyrosine kinase.

PMID:11287610

GAB2 is also a signaling protein, which is involved in the following pathway(s).

Scaffold Protein(s)	Related Pathway(s)
SHC1	MAPK1->LCK->ZAP70->GAB2
LAT	LCK->ZAP70->GAB2
CD247	LCK->ZAP70->GAB2
PTPN6	LCK->ZAP70->GAB2

The sequence of GAB2

Amino acids in red are phosphorylation sites determined by MS/MS

MSGGGDVVCT GWLRKSPPEK KLRRYAWKKR WFILRSGRMS GDPDVLEYYK 50
NDHSKKPLRI INLNFCEQVD AGLTFNKKEL QDSFVFDIKT SERTFYLVAE 100
TEEDMNKWVQ SICQICGFNQ AEESTDSLRN VSSAGHGPRS SPAELSSSSQ 150
HLLRERKSSA PSHSSQPTLF TFEPPVSNHM QPTLSTSAPQ EYLYLHQCIS 200
RRAENARSAS FSQGTRASFL MRSDTAVQKL AQGNGHCVNG ISGQVHGFYS 250
LPKPSRHNTE FRDSTYDLPR SLASHGHTKG SLTGSETDNE DVYTFKTPSN 300
TLCREFGDLL VDNMDVPATP LSAYQIPRTF TLDKNHNAMT VATPGDSAIA 350
PPPRPPKPSQ AETPRWGSPQ QRPPISENSR SVAATIPRRN TLPAMDNSRL 400
HRASSCETYE YPQRGGESAG RSAESMSDGV GSFLPGKMIV GRSDSTNSED 450
NYVPMNPGSS TLLAMERAGD NSQSVYIPMS PGAHHFDSLG YPSTTLPVHR 500
GPSRGSEIQP PPVNRNLKPD RKAKPTPLDL RNNTVIDELP FKSPITKSWS 550
RANHTFNSSS SQYCRPISTQ SITSTDSGDS EENYVPMQNP VSASPVPSGT 600
NSPAPKKSTG SVDYLALDFQ PSSPSPHRKP STSSVTSDEK VDYVQVDKEK 650
TQALQNTMQE WTDVRQSSEP SKGAKL

Home

Fig. 3 The ScaPD example for protein GAB2. The proteins is a scaffold protein and also a kinase

corresponding information page for the input protein. If the protein is a scaffold protein, the page will list the associated signaling pathways. Since the scaffold proteins are likely to be regulated through phosphorylation [18], the known phosphorylation sites are highlighted in the protein sequence. If the input protein is a signaling protein, the page will list the scaffold proteins that are associated with the pathways which the input protein is involved. Note that the protein names in the return pages are all clickable so that the users can navigate through the scaffold protein-mediated signaling pathways. In addition, we also provide the reference(s) that described the scaffold protein of interest (Fig. 3).

Discussion and Conclusion

Recent studies have revealed that the scaffold proteins play a versatile and important role in many signaling pathways. However, only a few scaffold proteins have been extensively characterized. Furthermore, no database has been developed for analyzing scaffold proteins, although many databases exist for signaling pathways. To our knowledge, ScaPD is the most comprehensive database focused on the scaffold proteins and associated signaling pathways. It holds a significant number of predicted scaffold proteins and their associated signaling pathways, which were previously completely uncharacterized. In addition, the database is more than a list of scaffold proteins. The users can search for scaffold proteins or singlaing pathways and their associated scaffold proteins. We will continuously update the scaffold proteins as new data are brought forth. Therefore, the ScaPD should provide additional information on the function of the scaffold proteins and pathways in signal transduction.

Author details
[1]Department of Ophthalmology, Johns Hopkins School of Medicine, Baltimore, MD, USA. [2]Department of Gastroenterology, Zhejiang Provincial People's Hospital, Hangzhou, Zhejiang, China. [3]The Horace Mann School in Bronx, Bronx, NY, USA. [4]Department of Pharmacology and Molecular Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA. [5]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, USA. [6]Center for High-Throughput Biology, Johns Hopkins School of Medicine, Baltimore, MD, USA.

References
1. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers–the database of key numbers in molecular and cell biology. Nucleic Acids Res. 2010;38(Database issue):D750–3. https://doi.org/10.1093/nar/gkp889.
2. Choi KY, Satterberg B, Lyons DM, Elion EA. Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in S. Cerevisiae. Cell. 1994;78(3):499–512.
3. Kornau HC, Schenker LT, Kennedy MB, Seeburg PH. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science. 1995;269(5231):1737–40.
4. Printen JA, Sprague GF Jr. Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade. Genetics. 1994;138(3):609–19.
5. Therrien M, Morrison DK, Wong AM, Rubin GM. A genetic screen for modifiers of a kinase suppressor of Ras-dependent rough eye phenotype in drosophila. Genetics. 2000;156(3):1231–42.
6. Zhang W, Sloan-Lancaster J, Kitchen J, Trible RP, Samelson LE. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. Cell. 1998;92(1):83–92.
7. Tsunoda S, Sierralta J, Sun Y, Bodner R, Suzuki E, Becker A, et al. A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. Nature. 1997;388(6639):243–9. doi:10.1038/40805.
8. McKay MM, Ritt DA, Morrison DK. Signaling dynamics of the KSR1 scaffold complex. Proc Natl Acad Sci U S A. 2009;106(27):11022–7. doi:10.1073/pnas.0901590106.
9. Strickfaden SC, Winters MJ, Ben-Ari G, Lamson RE, Tyers M, Pryciak PM. A mechanism for cell-cycle regulation of MAP kinase signaling in a yeast differentiation pathway. Cell. 2007;128(3):519–31. doi:10.1016/j.cell.2006.12.032.
10. Bhattacharyya RP, Remenyi A, Good MC, Bashor CJ, Falick AM, Lim WA. The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. Science. 2006;311(5762):822–6. doi:10.1126/science.1120941.
11. Malleshaiah MK, Shahrezaei V, Swain PS, Michnick SW. The scaffold protein Ste5 directly controls a switch-like mating decision in yeast. Nature. 2010;465(7294):101–5. doi:10.1038/nature08946.
12. Shaw AS, Filbert EL. Scaffold proteins and immune-cell signalling. Nat Rev Immunol. 2009;9(1):47–56. doi:10.1038/nri2473.
13. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. Annu Rev Biochem. 2006;75:655–80. doi:10.1146/annurev.biochem.75.103004.142710.

14. Good MC, Zalatan JG, Lim WA. Scaffold proteins: hubs for controlling the flow of cellular information. Science. 2011;332(6030):680–6. doi:10.1126/science.1198701.

15. Hu J, Rho HS, Newman RH, Zhang J, Zhu H, Qian J. PhosphoNetworks: a database for human phosphorylation networks. Bioinformatics. 2014;30(1): 141–2. doi:10.1093/bioinformatics/btt627.

16. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

17. Sadowski I, Breitkreutz BJ, Stark C, Su TC, Dahabieh M, Raithatha S, et al. The PhosphoGRID Saccharomyces Cerevisiae protein phosphorylation site database: version 2.0 update. Database. 2013;2013:bat026. doi:10.1093/database/bat026.

18. Hu J, Neiswinger J, Zhang J, Zhu H, Qian J. Systematic prediction of scaffold proteins reveals new design principles in scaffold-mediated signal transduction. PLoS Comput Biol. 2015;11(9):e1004508. doi:10.1371/journal.pcbi.1004508.

19. Morrison DK, Davis RJ. Regulation of MAP kinase signaling modules by scaffold proteins in mammals. Annu Rev Cell Dev Biol. 2003;19:91–118. doi:10.1146/annurev.cellbio.19.111401.091942.